# The Hitchhiker's Guide to the Social Media Data Research Galaxy - A Primer

Arianna Rossi(✉)

SnT, University of Luxembourg, Esch-sur-Alzette, Luxembourg
arianna.rossi@uni.lu

**Abstract.** This short paper is a primer for early career researchers that collect and analyze social media data. It provides concise, practical instructions on how to address personal data protection concerns and implement research ethics principles.

**Keywords:** Privacy and Data Protection · Research Ethics · Social Media Data

## 1 Introduction

We often hear students and colleagues deplore that the legal and ethical principles concerning Information and Communication Technologies (ICT) research are too abstract to be easily applied and that hands-on guidelines would enable them to carry out their work more efficiently and more respectfully. When it comes to internet-mediated research, namely to the social media "data gold mine" [11], its richness and availability may lure researchers into drawing heavily from it. However, harvesting social media data for research purposes raises ethical concerns, as well as questions about its legitimacy and lawfulness: being able to easily access such data does not imply that everything is allowed.

This article provides essential pointers about data protection and research ethics to early career scientists when they mine social media data, i.e., when they use automated means to index, analyse, evaluate and interpret mass quantities of content and data [14]. They do not consist in a checklist that can be blindly and effortlessly followed, but they are rather meant to raise questions and reflections to which researchers should find thoughtful answers. Both (personal) data management and ethical practices are generally (i) reflective: scientists need to reflect on their own specific research practices and the associated risks; (ii) dialogical: scientists must continuously review their practices in light of new experiences and new knowledge [13].

We have described at length our own approach to social media data management and research ethics in [24]. This article simplifies what we have discovered about social media data mining and summarizes it in an accessible, jargon-free language directed to early career researchers without a background in law or


F. Bieker et al. (Eds.): Privacy and Identity 2022, IFIP AICT 671, pp. 52–65, 2023.
https://doi.org/10.1007/978-3-031-31971-6_6

ethics. We organize the topics in two main sections, i.e., Sect. 3 on personal data protection and Sect. 4 on research ethics, even though there is reciprocal influence on the decisions and measures that should be adopted, for instance because protecting the confidentiality of data is a fundamental ethical principle.

## 2 Social Media Data

### 2.1 What Kind of Data Do We Collect?

Our first concern should be about the type of data that we mine: data collected on the internet can consist of text, images, videos, audio, code, etc. They can be collected from websites or concern user's navigation data; from datasets in the public domain or that are protected by specific access and reuse conditions; they can be recorded by sensors, like the heartbeat or voice and video recording; and many more. This article focuses on a specific type of internet data: the tremendous quantity of information produced within social media platforms, consisting of text (e.g., opinions, comments, preferences, links, etc.), images (e.g., pictures, GIFs, etc.), videos, users' social networks, metadata like timestamps and location markers of publication, among the others.

### 2.2 From Which Social Media Do We Collect Data?

We may ask which platform is the most appropriate to study the phenomenon we intend to investigate. The major social networks provide API to allow developers to access their content, including Twitter, Reddit, YouTube, Discord, Instagram, Facebook, LinkedIn, TikTok, Tumblr.

*Tip:* We are usually required to register on the platform as developers, disclose our data mining intentions and obtain their approval. Outside such controlled settings, web scraping is generally prohibited by the terms and conditions of the platforms, even though Article 3 of the Directive (EU) 2019/790 (also known as the Directive on Copyright in the Digital Single Market) allows scientific institutions to perform text and data mining for non-commercial purposes [8].

## 3 The Protection of Personal Data

The General Data Protection Regulation (GDPR) is the main piece of legislation that protects personal information in the EU[1]. The GDPR is based on the principles of transparency, lawfulness and accountability. This section will also touch upon data minimisation, purpose limitation, accuracy, storage limitation and integrity and confidentiality.

[1] Text available at: https://gdpr-info.eu/.

### 3.1 Do We Collect Personal Data?

The GDPR defines personal data as "any information that relates to an identified or identifiable living individual" (Article 4). It is not always straightforward to determine whether certain pieces of information, when combined, can lead to the identification of a person: on social networks, personal data encompass usernames and pseudonyms, pictures and videos, location markers and timestamps, mentions of other users, shared user networks, etc.

*Tip:* We should not only consider individual data points, but their combinations. The support of the Data Protection Officer (DPO) of the university can help us determine whether we collect personal data and which ones.

### 3.2 Can We Fully Anonymize Personal Data?

It is very challenging to effectively anonymize data so that it becomes impossible to identify the person to whom they refer. For example, a simple search online of a social media post can retrieve its original version and its author [5,20]. Moreover, de-anonymization techniques can revert the process of anonymization [29]. Lastly, effective anonymization may lower the utility of the data [3,19].

*Tip:* There are dedicated applications available online for anonymizing certain types of data[2]. However, merely masking identifiers (e.g., by replacing a personal name with a code) is not an anonymization technique, but rather a pseudonymization technique. In such cases, the data are still considered personal data that can be stored, if security measures and other data protection mechanisms are in place.

### 3.3 What Data Do We Need and Why?

A good research data management plan is based on a prior reflection on what data are actually needed, adequate and relevant to the purpose, how to collect them, as well as when and how to dispose of them, with the overall goal of minimizing data collection and retention (data minimization principle (Article 5(1)(c))). The purpose must be specified and the (further) processing or re-use of data must be compatible with such purpose [7] (e.g., data that were initially collected for scientific research cannot be repurposed for marketing).

*Tip:* Although it is sometimes difficult to strictly determine the types of data that will be necessary for the analysis before it starts, defining the purpose as specifically as possible and limiting data collection and storage to the strictly necessary has the benefit of lowering the risks that we need to mitigate. As a consequence, it we can simplify the data management measures we need to put in place.

[2] E.g., https://arx.deidentifier.org/ and https://bitbucket.org/ukda/ukds.tools.textanonhelper/wiki/browse/.

### 3.4 On Which Ground Do We Collect Data?

Personal data processing needs to be justified by a legal basis. For research purposes, the legal basis may be consent, public interest or legitimate interest (Article 6) [7], and may be additionally regulated by national law.

*Tip:* It is recommended to seek guidance from the DPO to determine the most appropriate legal basis. If this is consent, only under specific conditions it is lawful: it must be freely given, specific, informed and unambiguous (Article 4(11) and Article 7); additionally, it must be explicit if the data are of sensitive nature (e.g., health data, religious beliefs, etc. see Article 9 GDPR). Note that no matter the legal basis, a different ("ethical") consent to participation to scientific research must be obtained (see Sec. 4.2).

### 3.5 How Risky and What Kind of Impact Has Our Research on the Concerned Individuals?

When our activities are likely to result in high risks for the rights and freedoms of the individuals to whom the personal data refer (Article 35(1)), a Data Protection Impact Assessment (DPIA) becomes necessary [1]. The DPIA offers a structured way to assess and mitigate the risks entailed in data processing [15], when for example this implicates a systematic, extensive and automated evaluation (e.g., profiling) and when sensitive data are processed on a large scale.

*Tip:* Consulting with the DPO will clarify the need for a DPIA and the procedures. The project "Privacy in Research: Asking the Right Questions"[3] offers a dedicated online course, a roleplay game and a booklet [15], while official EU guidelines provide practical guidance questions [9].

### 3.6 How Do We Ensure the Confidentiality of the Data?

We can protect the confidentiality of personal data against unauthorized access via various (organizational and technical) security measures, for example encryption that irreversibly converts the data into an unreadable form, unless for those having the decryption key [27]; storage on the institutional GitLab and internally hosted servers (versus external clouds); strong authentication procedures; access control measures [27] that authorize or prohibit certain actions depending on the user role (e.g., students vs senior researchers); reversible pseudonymization that removes certain identifiers (multiple approaches exist, see [10]) and allows for the re-identification of users and for their consent gathering and withdrawal (see also Sec. 4.2).

*Tip:* The university often provides applications for anonymizing, pseudonymizing, encrypting and securely storing, transferring and sharing the data. Check the institutional IT catalogue or ask the DPO and the technical service of your institution. For a non-exhaustive list of open source tools for pseudonymization, refer to [25].

[3] https://sites.google.com/rug.nl/privacy-in-research/home?authuser=0.

### 3.7 How Do We Limit the Storage of the Data?

We can keep personal data only as long as necessary [21], then they should be deleted or anonymized. However, institutional and funding policies often impose to keep the data for longer periods of time out of data integrity and transparency reasons. This is why an exception applies to scientific research [7], provided that it is justified and security mechanisms are implemented [17].

*Tip:* Check institutional and funding policies to find out the required storage period and dispose mechanisms for the subsequent data deletion.

### 3.8 How Do We Ensure the Transparency of Our Data Practices?

Before the data collection starts, we need to inform the concerned people about what we do with their personal data, why we are authorized to gather them, how we are going to use them, and about their rights (Articles 13–14). When we indirectly collect personal data from a social media platform, rather than directly from the concerned individuals, and when informing thousands would constitute a disproportionate effort [7,17], it is admissible to draft a public notice addressed to the people that may be impacted by the research study. The notice must be easily accessible and easy to understand (Article 12).

*Tip:* The DPO or the Institutional Review Board (IRB) usually provide standard privacy notices that must be tailored to the specific case and can be published on the website of the research project. To be sure everybody understands, we should use plain language and a conversational style, and clarify technical jargon; to allow for skim reading, we should highlight the main words and structure the document in sections with informative headings [2,23]. Moreover, we can devise a short notice [26] that is periodically published on the social media (e.g., a tweet) to warn users of the (imperceptible) data collection and of the opt-out option detailed in Sect. 4.2, with a link to the full public notice on the website.

### 3.9 How Do We Enable the Exercise of Rights?

Data protection envisages several rights for the individuals to whom data refer: the right to access (Art. 15), rectify (Art. 16), erase (Art. 17) and transfer (Art. 17) their data, as well as to ask to limit (Art. 18) and object to processing (Art. 21), which would prevent us from using the data. If the legal basis is consent, we need to provide an easy way to withdraw consent (Art. 7(3)). Individuals can also submit complaints to the relevant data protection authority (Art. 77), but this should be the last resort: it is advisable to seek to solve all controversies beforehand.

*Tip:* We need to establish appropriate measures for people to act upon their rights. For example, describe how exactly we enable them to access their data and put in place user-friendly mechanisms to do so.

### 3.10 Useful Resources on Data Protection Compliance

Additional guidance with examples is provided in [21] and the other sources cited throughout this section. Institutions and national research bodies usually provide resources as well. In the Appendix A, we provide a diagram that guides researchers through the process.

## 4 Research Ethics in Practice

Similarly to data protection approaches, research ethics decisions are based on judgement calls [13]: what constitutes ethical behaviour depends on the context where the study occurs [30] and is guided by the researchers asking the right questions. Thus, ambiguity, uncertainty and disagreement are inevitable [13], especially to solve ethical dilemmas on internet-mediated data where best practices may not be established. Valuing interdisciplinary knowledge and collective problem-solving is a helpful way to answer such dilemmas, as well as resorting to the organizational and sectorial code of conduct. Further, in many universities, the ethical authorization of the Institutional Review Board should be sought before starting a study.

### 4.1 Research on Social Media Data Equals Research with Human Subjects

In Europe, the collection and analysis of social media data is considered research with human subjects and must therefore abide by the same ethical safeguards [16]. The difference, however, is that the authors of online contents are largely unaware that their input may be collected and reused elsewhere for scientific purposes [12,22], unless they are directly told so. Thus, their participation to our research study cannot be considered voluntary nor informed, which raises novel challenges.

### 4.2 How May We Respect the Autonomy, Privacy, and Dignity of Individuals and Communities?

**Public Vs. Private Information.** People disclose information on social media platforms even on sensitive, intimate topic to feel like they belong to a community and share their experiences. Despite its availability and seemingly public nature, we should not lightheartedly scrape, analyze, or report such information in other context. We may infer what constitutes acceptable behaviour from contextual elements. For instance, content disclosed on a password-protected group is likely more private [16] than content published with a visible and retrievable hashtag that contributes to a public debate. Moreover, when the information is sensitive (e.g., sexual orientation, health-related issues, etc.) or concerns minors, it may cause harm if it is divulged outside of its intended context.

*Tip:* We should ask whether the scientific benefits of the disclosure to a different audience in a different context outweigh the possible harms [16]. The answer may also be negative. Depending on the cases, the permission of the authors, of the group moderator or of the legal guardian should be sought [12,31].

**Anonymity and Confidentiality.** By implementing confidentiality measures, we mitigate the risk of increased vulnerability, embarrassment, reputation damage, or prosecution of the subjects of the study [28]. Since public social media posts are retrievable via a simple online search, when we republish e.g., verbatim quotes in articles, it is difficult to conceal the authors' identities.

*Tip:* We should not quote verbatim contents without a good reason [12] or ask for the authors' permission (this is where pseudonymization can be handy). Paraphrasing the content can constitute a viable solution [16] in this respect, however it may impact data accuracy and integrity (Sec. 4.3).

**Informed Consent and Withdrawal.** Based on the disclosure of the practices and risks involved in the study, informed consent allows individuals to freely express their willingness to join or abstain as research participants, under the condition that they should be able to withdraw their consent whenever they desire. In large-scale data mining, seeking consent from each potential participant may be impracticable [13], though. In such cases and when the collection concerns non-sensitive, visibly public data (see Sect. 4.2) [16], providing transparency about the research study [12] (see also Sec. 3.8) and the possibility to withdraw (i.e., retrospective withdrawal [16] or opt-out) may be considered sufficient.

*Tip:* The consent to research participation ("ethical consent") is always required and differs from the ("legal") consent that may or may not constitute the legal basis for personal data processing (see Sect. 3.4). Whenever possible, ethical consent for data scraping from research participants should be sought. Otherwise, we can aggregate data and ask consent for republication of individual quotes or images. If the information is sensitive and not visibly public, it is advisable to ask the permission to use it to the group moderator, at least. We should also set up procedures to allow impacted users to contact us to opt-out from the study and delete or exclude their data from any further processing.

### 4.3 Scientific Integrity

This principle concerns the quality, integrity and contribution of a piece of research [16].

**Data Quality.** Data should not be tampered nor transformed in an unreasonable manner to ensure the validity of the research findings and conclusions [16]. Moreover, data coming from social networks may be produced by bots and exclude those voices and experiences that are absent from the platform. Further, we do not know whether the data made available via the platform's APIs is representative and complete. Lastly, inaccurate and even misleading inferences [16] may be generated through data analysis.

*Tip*: We should be aware of and mitigate such risks, starting with a transparent account of such shortcomings. We should also resort to an anticipatory assessment of the entailed risks and balance them against the envisaged benefits deriving from the research study. When the risks outweigh the benefits and mitigation measures do not suffice, we need to find alternative ways to get to the scientific answers we seek.

**Minors.** As the minimal age requirement to register on many social media is 13, we may unwittingly involve minors and their data in our study without the necessary authorization of their legal guardian [16].

*Tip:* At present, there seems to be no reliable manner to exclude minors' data from data collection. Thus we should report the results only in an aggregate form and paraphrase quoted if we intend to republish them.

### 4.4 Social Responsibility

This principle enshrines that researchers should act in the public interest by respecting social structures of (online) communities, avoiding their disruption, and thoughtfully balancing the positive and negative consequences of a piece of research [16].

**Invasive Research and Dual Use.** Building on the tension between public and private domain (Sec. 4.2), there may be certain communities that do not welcome being under the spotlight nor analyses of their disclosures in a trustworthy online environment. Stretching the meaning of dual use, it is paramount to reflect in an anticipatory manner on whether the data, code or knowledge that we produce can be used to harm the society (e.g., job loss, discrimination), the environment (e.g., pollution, global warming), living beings' health and safety (e.g., military applications), and the rights and freedoms (e.g., privacy) [4]. We should also propose measures to mitigate such risks or even avoid certain research activities if deemed too risky.

*Tip:* We should apply the proportionality principle which requires that ethical checks and procedures should be proportional to the assessed risks and potential harms [16] and weight the benefits of the findings against the harms in each research context, as part of a research impact assessment.

**Risks for Researchers.** Using social media can expose researchers to cyber-risks (e.g., cyberbulling, stalking and media pillories) as well as physical threats, which can lead to professional and personal consequences like loss of scientific reputation and anxiety.

*Tip:* Such risks must be mitigated, especially when they concern early stage researchers, mainly through institutional preparedness [18].

**Compensation.** It is common practice to reward research participants for their time and efforts, but doing so with data that was collected indirectly is cumbersome or even impossible.

*Tip:* We may reciprocate with enhanced knowledge based on the publication of the study results [12].

### 4.5 Maximise Benefits and Minimise Harm

Confidentiality and security measures, precautions against using minors' data, processes for consent collection and withdrawal, as well as oversight over the ethicality of the process and the validity of the research, are meant to minimize the risks for participants and researchers, while maximizing the research benefits [16].

### 4.6 Useful Resources on Research Ethics

Additional guidance with examples is provided in [6,13,16] as well as in professional code of conducts, like the ACM code[4].

## 5 Limitations

The content of this article is meant to be a primer and must be taken critically as each research study has its own features and what applies to one case may not apply to seemingly similar ones. Since it oversimplifies many topics, we invite the readers to explore the proposed bibliography resources and beyond. Moreover, this short guide has not been tested in use and is not complete: for example, sustainability as an ICT ethical principle is not mentioned. Moreover, researchers should get acquainted with FAIR principles (i.e., findable, accessible,

[4] https://www.acm.org/code-of-ethics.

interoperable, reusable)[5] for data management and be able to choose the appropriate licenses for software[6] and data[7]. To get a broader understanding of such issues, DCC provides how-to guides and checklists on their website[8], whereas CESSDA has published a complete guide to data management online[9] that can be learnt through a serious game[10].

## 6 Conclusions

In a nutshell, the main points to retain are:

1. Start planning the data management process and contact the DPO and the IRB well before you start the data collection as it may take time to receive an authorization;
2. Ask for support from more experienced researchers and to those with complementary experiences;
3. Check the institutional resources (e.g., code of ethics; consent templates; security-enhancing tools; data protection training; etc.) available to you;
4. It is safer to assume that social media data and metadata are personal data, rather than the contrary;
5. Analysing information published online counts as research on human subjects, thus the same ethical safeguards should be upheld;
6. Thus, ethical consent to research participation should be sought; under certain conditions, transparency and opt-out may suffice;
7. Actual anonymization is hard to achieve, while pseudonymization allows for re-identification (i.e., to contact the users, enable them to opt-out, uphold data integrity) but other technical and organizational measures are needed.

**Acknowledgements.** This work has been partially supported by the Luxembourg National Research Fund (FNR) - IS/14717072 "Deceptive Patterns Online (Decepticon)" and the H2020-EU grant agreement ID 956562 "Legally-Attentive Data Scientists (LeADS)". The main content of this work was published in extend form in [24] and re-elaborated thanks to the comments of colleagues, students and data management experts.

[5] https://www.go-fair.org/fair-principles/.
[6] https://choosealicense.com/.
[7] https://ufal.github.io/public-license-selector/.
[8] https://www.dcc.ac.uk/guidance/how-guides.
[9] https://dmeg.cessda.eu/Data-Management-Expert-Guide.
[10] https://lod.sshopencloud.eu/.

# A Appendix A

See the Fig. 1.

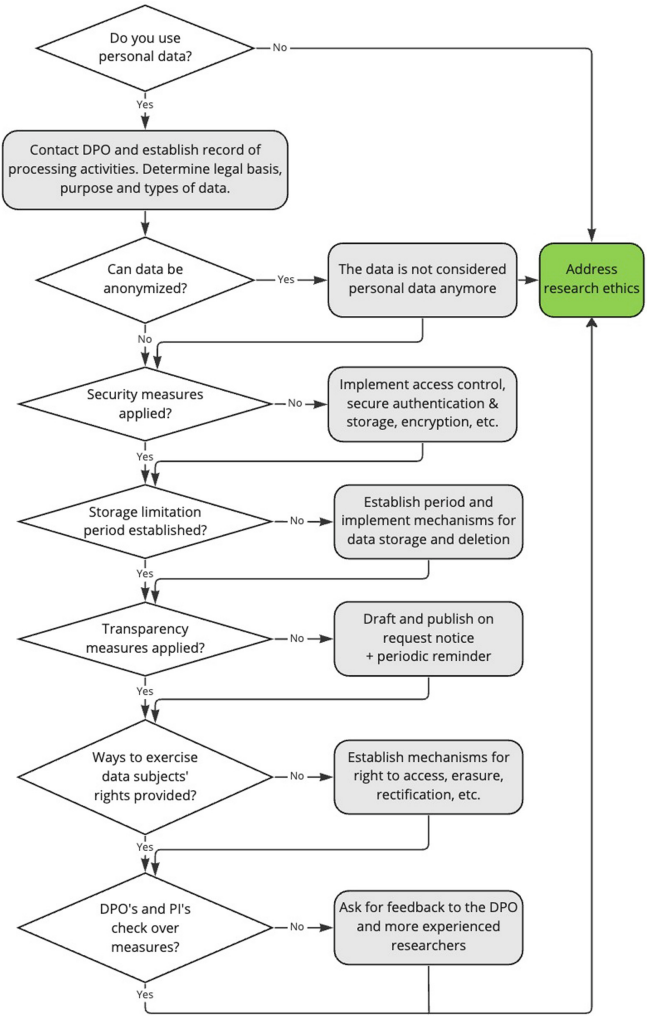


**Fig. 1.** Decision diagram for data protection measures

## References

1. Article 29 Data Protection Working Party: Guidelines on data protection impact assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of regulation 2016/679 (2017). https://ec.europa.eu/newsroom/just/document.cfm?doc_id=47711

2. Article 29 Data Protection Working Party: Guidelines on transparency under regulation 2016/679, 17/EN WP260 rev.01 (2018–04). https://ec.europa.eu/newsroom/article29/redirection/document/51025
3. Article 29 Working Party: Opinion 05/2014 on anonymisation techniques (2014). https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
4. Benčin, R., Strle, G.: Social responsibility in science and engineering (2015–06). https://satoriproject.eu/media/1.c-Social-responsibility.pdf
5. Beninger, K., Fry, A., Jago, N., Lepps, H., Nass, L., Silvester, H.: Research using social media; users' views (2014). https://www.natcen.ac.uk/media/282288/p0639-research-using-social-media-report-final-190214.pdf
6. Commission, E.: Horizon 2020 guidance - how to complete your ethics self-assessment (2018). https://eneri.eu/wp-content/uploads/2018/10/H2020-Guidance-How-to-complete-your-ethics-self-assessment.pdf
7. Ducato, R.: Data protection, scientific research, and the role of information. Comput. Law Secur. Rev. **37**, 105412 (2020). https://doi.org/10.1016/j.clsr.2020.105412, https://www.sciencedirect.com/science/article/pii/S0267364920300170
8. Ducato, R., Strowel, A.M.: Ensuring text and data mining: remaining issues with the EU copyright exceptions and possible ways out. Eur. Intell. Prop. Rev. **43**(5), 322–337 (2021). https://papers.ssrn.com/abstract=3829858
9. European Data Protection Supervisor: Accountability on the ground part II: Data protection impact assessments & prior consultation (2018). https://edps.europa.eu/sites/edp/files/publication/18-02-06_accountability_on_the_ground_part_2_en.pdf
10. European Network and Information Security Agency: Pseudonymisation techniques and best practices: recommendations on shaping technology according to data protection and privacy provisions. Publications Office (2019). https://data.europa.eu/doi/10.2824/247711
11. Fiesler, C., Beard, N., Keegan, B.C.: No robots, spiders, or scrapers: Legal and ethical regulation of data collection methods in social media terms of service. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 14, pp. 187–196 (2020). https://ojs.aaai.org/index.php/ICWSM/article/view/7290
12. Fiesler, C., Proferes, N.: "participant" perceptions of twitter research ethics. Soc. Media + Soc. **4**(1), 2056305118763366 (2018). https://doi.org/10.1177/2056305118763366
13. Franzke, A.S., Bechmann, Anja, Zimmer, M., Ess, Charles M.: Internet research: Ethical guidelines 3.0: Association of internet researchers (2019). https://aoir.org/reports/ethics3.pdf
14. Future TDM: D3.3+ baseline report of policies and barriers of TDM in europe (2016). https://project.futuretdm.eu/wp-content/uploads/2017/05/FutureTDM_D3.3-Baseline-Report-of-Policies-and-Barriers-of-TDM-in-Europe.pdf
15. Hoorn, E., Montagner, C.: Starting with a DPIA methodology for human subject research (2018). https://www.rug.nl/research/research-data-management/downloads/c2-dataprotection-dl/dpia_guidance_doc_v1_pub.pdf
16. Kaye, L.K., et al.: Ethics guidelines for internet-mediated research (2021). https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy%20-%20Files/Ethics%20Guidelines%20for%20Internet-mediated%20Research.pdf. ISBN: 978-1-85433-796-2

17. Kuyumdzhieva, A.: General data protection regulation and horizon 2020 ethics review process: ethics compliance under GDPR. Bioethica. **5**(1), 6–12 (2019). https://doi.org/10.12681/bioeth.20832, https://ejournals.epublishing.ekt.gr/index.php/bioethica/article/view/20832
18. Marwick, A.E., Blackwell, L., Lo, K.: Best practices for conducting risky research and protecting yourself from online harassment. In: Data & Society Guide, p. 10. Data & Society Research Institute (2016). https://datasociety.net/wp-content/uploads/2016/10/Best_Practices_for_Conducting_Risky_Research-Oct-2016.pdf
19. Mészáros, J., Ho, C.h.: Big data and scientific research: the secondary use of personal data under the research exemption in the GDPR. Hungarian J. Legal Stud. **59**(4), 403–419 (2018). https://doi.org/10.1556/2052.2018.59.4.5, https://akjournals.com/doi/10.1556/2052.2018.59.4.5
20. Narayanan, A., Shmatikov, V.: De-anonymizing social networks. In: 2009 30th IEEE Symposium on Security and Privacy, pp. 173–187 (2009). https://doi.org/10.1109/SP.2009.22, ISSN: 2375-1207
21. Officer, I.C.: Guide to the UK general data protection regulation (UK GDPR). https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/
22. Proferes, N.: Information flow solipsism in an exploratory study of beliefs about twitter. Soc. Media + Soc. **3**(1), 2056305117698493 (2017). https://doi.org/10.1177/2056305117698493
23. Rossi, A., Ducato, R., Haapio, H., Passera, S.: When design met law: design patterns for information transparency. Droit de la Consommation = Consumenterecht : DCCR. **122**, 79–121 (2019). https://orbilu.uni.lu/bitstream/10993/40116/1/A.%20Rossi,%20R.%20Ducato,%20H.%20Haapio%20et%20S.%20Passera.pdf
24. Rossi, A., Kumari, A., Lenzini, G.: Unwinding a legal and ethical ariadne's thread out of the twitter's scraping maze. In: Privacy Symposium 2022 - Data Protection Law International Convergence and Compliance with Innovative Technologies (DPLICIT). Springer Nature (2022). https://doi.org/10.1007/978-3-031-09901-4_10
25. Rossi, A., Arena, M.P., Kocyijit, E., Hani, M.: Challenges of protecting confidentiality in social media data and their ethical import. In: 1st International Workshop on Ethics in Computer Security (EthiCS 2022) Co-located with the 7th IEEE European Symposium on Security and Privacy, pp. 554–561. IEEE (2022). https://doi.org/10.1109/EuroSPW55150.2022.00066
26. Schaub, F., Balebako, R., Durity, A.L., Cranor, L.F.: A design space for effective privacy notices. In: Proceedings of Eleventh Symposium On Usable Privacy and Security (SOUPS 2015), pp. 1–17 (2015). https://doi.org/10.5555/3235866.3235868
27. Stallings, W.: Operating system security (chapter 24). In: Computer Security Handbook, pp. 24.1–24.21. Wiley, 6 edn. (2014). https://doi.org/10.1002/9781118851678.ch24
28. Townsend, L., Wallace, C.: The ethics of using social media data in research: a new framework. In: Woodfield, K. (ed.) The Ethics of Online Research, Advances in Research Ethics and Integrity, vol. 2, pp. 189–207. Emerald Publishing Limited (2017). https://doi.org/10.1108/S2398-601820180000002008
29. Tripathy, B.K.: De-anonymization techniques for social networks. In: Dey, N., Borah, S., Babo, R., Ashour, A.S. (eds.) Social Network Analytics, pp. 71–85. Academic Press (2019). https://doi.org/10.1016/B978-0-12-815458-8.00004-9, https://www.sciencedirect.com/science/article/pii/B9780128154588000049

30. Vitak, J., Shilton, K., Ashktorab, Z.: Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, pp. 941–953. ACM (2016). https://doi.org/10.1145/2818048.2820078
31. Williams, M.L., Burnap, P., Sloan, L.: Towards an ethical framework for publishing twitter data in social research: Taking into account users' views, online context and algorithmic estimation. Sociology **51**(6), 1149–1168 (2017). https://doi.org/10.1177/0038038517708140